Effects of data dimension on empirical likelihood

Author(s): SONG XI CHEN, LIANG PENG and YING-LI QIN

Source: *Biometrika*, Vol. 96, No. 3 (SEPTEMBER 2009), pp. 711-722
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/27798858
Accessed: 24-05-2021 08:40 UTC

Biometrika (2009), **96**, 3, *pp.* 711–722
doi: 10.1093/biomet/asp037

Printed in Great Britain
Advance Access publication 8 August 2009

Effects of data dimension on empirical likelihood

BY SONG XI CHEN
Department of Statistics, Iowa State University, Iowa 50011-1210, U.S.A.
songchen@iastate.edu

LIANG PENG
School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332-0160, U.S.A.
peng@math.gatech.edu

AND YING-LI QIN
Department of Statistics, Iowa State University, Iowa 50011-1210, U.S.A.
qinyl@iastate.edu

SUMMARY

We evaluate the effects of data dimension on the asymptotic normality of the empirical likelihood ratio for high-dimensional data under a general multivariate model. Data dimension and dependence among components of the multivariate random vector affect the empirical likelihood directly through the trace and the eigenvalues of the covariance matrix. The growth rates to infinity we obtain for the data dimension improve the rates of Hjort et al. (2008).

Some key words: Asymptotic normality; Data dimension; Empirical likelihood; High-dimensional data.

1. INTRODUCTION

Since Owen (1988, 1990) introduced empirical likelihood, it has been extended to a wide range of settings as a tool for nonparametric and semiparametric inference. Its most attractive property is its permitting likelihood-like inference in nonparametric or semiparametric settings, largely due to its sharing two key properties with the conventional likelihood: Wilks' theorem and Bartlett correction (Hall & La Scala, 1990; DiCiccio et al., 1991; Chen & Cui, 2006).

High-dimensional data are increasingly common; for instance, in DNA and genetic microarray analysis, marketing research and financial applications. There is a rapidly expanding literature on multivariate analysis where the data dimension p depends on the sample size n and grows to infinity as $n \to \infty$; see, for example, Portnoy (1984, 1985) in the context of M-estimation, Bai & Saranadasa (1996) for two-sample test for means, Ledoit & Wolf (2002) and Schott (2005) for testing a specific covariance structure and Schott (2007) for tests with more than two samples.

Given the interest in both high-dimensional data and empirical likelihood, there is a need to evaluate the behaviour of the latter when the data dimension and the sample size increase simultaneously. In this paper, we evaluate the effects of the data dimension and dependence on the asymptotic normality of the empirical likelihood ratio statistic for the mean.

Let $X_1, \ldots, X_n$ be independent and identically distributed p-dimensional random vectors in R^p with mean vector $\mu = (\mu_1, \ldots, \mu_p)^{\mathrm{T}}$ and nonsingular variance matrix Σ. Let

$$L_n(\mu) = \sup \left(\Pi_{i=1}^n \pi_i : \pi_i \geqslant 0, \sum_{i=1}^n \pi_i = 1, \sum_{i=1}^n \pi_i X_i = \mu \right) \tag{1}$$

be the empirical likelihood for μ and let $w_n(\mu) = -2\log\{n^n L_n(\mu)\}$ be the empirical likelihood ratio statistic. When p is fixed, Owen (1988, 1990) showed that

$$w_n(\mu) \to \chi_p^2 \tag{2}$$

in distribution as $n \to \infty$, which mimics Wilks' theorem for parametric likelihood ratios. An extension of the above result for parameters defined by general estimating equations is given in Qin & Lawless (1994).

As $p \to \infty$ for high-dimensional data, the natural substitute for (2) is

$$(2p)^{-1/2}\{w_n(\mu) - p\} \to N(0, 1) \tag{3}$$

in distribution as $n \to \infty$, since χ_p^2 is asymptotic normal with mean p and variance $2p$. A key question is how large the dimension p can be while (3) remains valid. In a recent study, Hjort et al. (2008) have established that it is $p = o(n^{1/3})$ under the assumptions:

Assumption 1. The eigenvalues of Σ are uniformly bounded away from zero and infinity, and

Assumption 2. All components of X_i are uniformly bounded random variables.

When Assumption 2 is relaxed, we have:

Assumption 2′. $E(||p^{-1/2}X_i||^q)$ and $p^{-1}\sum_{j=1}^p E(|X_i^{(j)} - \mu_j|^q)$ are bounded for some $q \geqslant 4$, where $||\cdot||$ is the Euclidean norm. Hjort et al. (2008) showed that (3) is valid if $p^{3+6/(q-2)}/n \to 0$.

When $q = 4$ in Assumption 2′, it means $p = o(n^{1/6})$. Hence, there is a significant slowing-down on the rate of $p \to \infty$ when Assumption 2 is weakened. Tsao (2004) found that, when p is moderately large but fixed, the distribution of $w_n(\mu)$ has an atom at infinity for fixed n: the probability of $w_n(\mu) = \infty$ is nonzero. Tsao showed that, if p and n increase at the same rate such that $p/n \geqslant 0{\cdot}5$, the probability of $w_n(\mu) = \infty$ converges to 1 since the probability of μ being contained in the convex hull of the sample converges to 0. These reveal the effects of p on the empirical likelihood from another perspective.

In this paper, we analyze the empirical likelihood for high-dimensional data under a general multivariate model, which facilitates a more detailed analysis than Hjort et al. (2008) and allows less restrictive conditions. The analysis requires neither the largest eigenvalue of Σ nor $E(||p^{-1/2}X_i||^q)$ to be bounded, and hence accommodates a wider range of dependences among components of X_i.

Our main finding is that the effect of the dimensionality and the dependence among components of X_i on the empirical likelihood are leveraged through $\mathrm{tr}(\Sigma)$, the trace of the covariance matrix Σ and its largest eigenvalue λ_p. We provide a general rate for the dimension p, which is shown to be dependent on $\mathrm{tr}(\Sigma)$ and λ_p. In particular, under Assumptions 1 and 2, $p = o(n^{1/2})$, which improves $p = o(n^{1/3})$ of Hjort et al. (2008). This is likely to be the best rate for p in the context of the empirical likelihood as $p = o(n^{1/2})$ is the sufficient and necessary condition for the convergence of the sample covariance matrix to Σ under the trace-norm when all the eigenvalues of Σ are bounded.

2. Preliminaries

Suppose that each of the independent and identically distributed observations $X_i \in R^p$ is specified by $X_i = \Gamma Z_i + \mu$, where Γ is a $p \times m$ matrix, $m \geqslant p$, and $Z_i = (Z_{i1}, \ldots, Z_{im})^{\mathrm{T}}$ is a random vector such that

$$E(Z_i) = 0, \mathrm{var}(Z_i) = I_m, \quad E\left(Z_{il}^{4k}\right) = m_{4k} \in (0, \infty), \qquad (4)$$
$$E\left(Z_{il_1}^{\alpha_1} \cdots Z_{il_q}^{\alpha_q}\right) = E\left(Z_{il_1}^{\alpha_1}\right) \cdots E\left(Z_{il_q}^{\alpha_q}\right),$$

whenever $\sum_{l=1}^{q} \alpha_l \leqslant 4k$ and $l_1 \neq \cdots \neq l_q$. Here k is some positive integer and I_m is the m-dimensional identity matrix.

The above multivariate model, employed in Bai & Saranadasa (1996), means that each X_i is a linear transformation of some m-variate random vector Z_i. An important feature is that m, the dimension of Z_i, is arbitrary provided $m \geqslant p$ and $\Gamma\Gamma^{\mathrm{T}} = \Sigma$, which can generate a rich collection of X_i from Z_i with the given covariance Σ. It also requires that power transformations of different components of Z_i are uncorrelated, which is weaker than assuming that they are independent. The model (4) encompasses many multivariate models. It includes the elliptically contoured distributions with $Z_i = RU^{(m)}$ where R is a nonnegative random variable and $U^{(m)}$ is the uniform random vector on the unit sphere (Fang & Zhang, 1990). The multivariate normal and t-distribution are elliptically contoured, and so are a mixture of normal distributions whose density is defined by $\int n(x \mid \mu, v^{-2}\Sigma) dw(v)$, where $n(x \mid \mu, \Sigma)$ is the density of $N(\mu, \Sigma)$ and $w(v)$ is the distribution function of a nonnegative univariate random variable (Anderson, 2003). Both the moment conditions and the correlation are imposed on Z_i rather than X_i. This model structure allows the moments of $||X_i - \mu||^{2k}$ to be derived and allows us to conduct a more detailed analysis than possible in Hjort et al. (2008).

The integer k determines the number of finite moments for Z_{il}. As $k \geqslant 1$, each Z_{il} has at least finite fourth moments. This is the minimal moment condition to ensure the convergence of the largest eigenvalue of the sample covariance matrix to the largest eigenvalues of Σ (Yin et al., 1988; Bai et al., 1998), and hence the convergence of the sample covariance matrix to Σ under the matrix norm based on the largest eigenvalue. By inspecting the proofs given in the Appendix, we see that a divergent sample covariance matrix would dramatically alter the asymptotic mean and variance of the empirical likelihood ratio. Hence, it is unclear if (3) would remain true.

From the standard empirical likelihood solutions (Owen, 1988, 1990) that are valid for any p, fixed or growing, the optimal weights π_i for the optimization problem (1) are

$$\pi_i = \frac{1}{n} \frac{1}{1 + \lambda^{\mathrm{T}}(X_i - \mu)},$$

where $\lambda \in R^p$ is a Lagrange multiplier satisfying

$$g(\lambda) = \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^{\mathrm{T}}(X_i - \mu)} = 0. \qquad (5)$$

Hence, the empirical likelihood $L_n(\mu)$ equals $n^{-n} \prod_{i=1}^{n} \{1 + \lambda^{\mathrm{T}}(X_i - \mu)\}^{-1}$. As the maximum empirical likelihood is attained at $\pi_i = n^{-1}$ $(i = 1, \ldots, n)$, the empirical likelihood ratio statistic is

$$w_n(\mu) = -2\log\{n^n L_n(\mu)\} = 2\sum_{i=1}^{n} \log\{1 + \lambda^{\mathrm{T}}(X_i - \mu)\}.$$

Throughout the paper we let $\gamma_1(A) \leqslant \cdots \leqslant \gamma_p(A)$ denote the eigenvalues and let $\mathrm{tr}(A)$ denote the trace operator of a matrix A. When $A = \Sigma$, we write $\gamma_j(\Sigma)$ as γ_j $(j = 1, \ldots, p)$. It is assumed throughout the paper that $\gamma_1 \geqslant C_1$ for some positive constant C_1.

3. EFFECTS OF HIGH DIMENSION

The Lagrange multiplier λ defined in (5) is a key element in any empirical likelihood formulation, and reflects the implicit nature of the methodology. When p is fixed, Owen (1990) showed that

$$||\lambda|| = O_p(n^{-1/2}). \tag{6}$$

This has been the prevailing order for the λ except in nonparametric curve estimation, where n is replaced by the effective sample size (Chen, 1996). When p grows with n, (6) is no longer valid.

THEOREM 1. *If* $\{\mathrm{tr}(\Sigma)\}^{4k-1}\gamma_p = O(n^{2k-1})$ *and* $\gamma_p^2 p^2 = o(n)$, *then* $||\lambda|| = O_p[\{\mathrm{tr}(\Sigma)/n\}^{1/2}]$.

Theorem 1 implies that the effect of the dimension and dependence among components of X_i on the Lagrange multiplier is directly determined through $\mathrm{tr}(\Sigma)$ and γ_p. The rate for $||\lambda||$ can be regarded as a generalization of (6) for a fixed p since $O_p[\{\mathrm{tr}(\Sigma)/n\}^{1/2}]$ degenerates to $O_p(n^{-1/2})$ in that case.

We first study the effects of dimension on the asymptotic normality of $w_n(\mu)$, assuming existence of the minimal fourth moment for each Z_{il}. Later, we will increase the number of moments. We assume for the time being that $k = 1$ in (4) and $\mathrm{tr}^5(\Sigma)\gamma_p^5 = o(np)$. Since $p\gamma_1 \leqslant \mathrm{tr}(\Sigma) \leqslant p\gamma_p$, this implies the conditions of Theorem 1.

We wish to establish an expansion for $w_n(\mu)$. Put $W_i = \lambda^{\mathrm{T}}(X_i - \mu)$. From (A7) of the Appendix, $\max_{i=1,\ldots,n} |W_i| = o_p(1)$, which allows

$$\log\{1 + \lambda^{\mathrm{T}}(X_i - \mu)\} = W_i - W_i^2/2 + W_i^3/(1 + \xi_{i1})^4, \tag{7}$$

where $|\xi_{i1}| \leqslant |\lambda^{\mathrm{T}}(X_i - \mu)|$. Expand (5) so that

$$0 = g(\lambda) = \bar{X} - \mu - S_n\lambda + \beta_n,$$

where $\beta_n = n^{-1}\sum_{i=1}^{n}(X_i - \mu)W_i^2/(1 + \xi_i)^3$ for some $|\xi_i| \leqslant |\lambda^{\mathrm{T}}(X_i - \mu)|$ and $S_n = n^{-1}\sum_{i=1}^{n}(X_i - \mu)(X_i - \mu)^{\mathrm{T}}$. Hence,

$$\lambda = S_n^{-1}(\bar{X} - \mu) + S_n^{-1}\beta_n. \tag{8}$$

From (7) and (8), we obtain an expansion for $w_n(\mu)$:

$$\begin{aligned} w_n(\mu) &= n(\bar{X} - \mu)^{\mathrm{T}} S_n^{-1}(\bar{X} - \mu) - n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}\textstyle\sum_{i=1}^{n}\{\lambda^{\mathrm{T}}(X_i - \mu)\}^3/(1 + \xi_i)^4 \\ &= n(\bar{X} - \mu)^{\mathrm{T}}\Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^{\mathrm{T}}(S_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) \\ &\quad -n\beta_n S_n^{-1}\beta_n + \tfrac{2}{3}R_n\{1 + o_p(1)\}, \end{aligned} \tag{9}$$

where $R_n = \sum_{i=1}^{n}\{\lambda^{\mathrm{T}}(X_i - \mu)\}^3$. This expansion looks similar to that given in Owen (1990) for a fixed p, but the stochastic order of each term requires careful evaluation as p grows with n.

From Lemma 5 in the Appendix, we have

$$(2p)^{-1/2}\{n(\bar{X}-\mu)^{\mathrm{T}}\Sigma^{-1}(\bar{X}-\mu)-p\} \to N(0,1) \tag{10}$$

in distribution as $n \to \infty$, which is true under much weaker conditions, for instance $p/n \to c \geqslant 0$, by applying the martingale central limit theorem. Derivations given in the Appendix show that the other two terms on the right-hand side of (9) are both $o_p(p^{1/2})$. These lead us to establish (3) as summarized in the following theorem.

THEOREM 2. *If* $k = 1$ *in* (4) *and* $\mathrm{tr}^5(\Sigma)\gamma_p^5 = o(np)$, *then* (3) *is valid.*

Theorem 2 indicates that, when γ_p is bounded, (3) is true if $p = o(n^{1/4})$, which improves the order $p = o(n^{1/6})$ obtained by Hjort et al. (2008) under the finite fourth moment condition of X_i, which we do not need in our study. The conditions assumed under Theorem 2 are liberal compared to Assumptions 1 and 2, and there is no explicit restriction on γ_p, which may diverge to infinity as $n \to \infty$.

Next we show that the dimension p can increase more rapidly if Z_{il} possesses more than four moments. Assuming higher-order moments allows us to evaluate those terms in (9) more accurately. Specifically, we will assume Z_{il} has at least finite 12th moment, $k \geqslant 3$ in model (4). The case $k \geqslant 2$ can be considered as a part of the case $k \geqslant 1$ whose analysis is covered by Theorem 2. The following theorem, whose proof is given in an Iowa State University technical report available from the authors, shows that $p = o(n^{1/2})$ is approachable.

THEOREM 3. *If* $k \geqslant 3$ *in* (4), $\{\mathrm{tr}(\Sigma)\}^{4k-1}\gamma_p = O(n^{2k-1})$ *and* $p^2\gamma_p^5 = o\{n^{(4k-1)/(4k)}\}$, *then* (3) *is valid.*

When γ_p is bounded, Theorem 3 implies that $w_n(\mu)$ is asymptotically normally distributed if $p = o(n^{1/2-1/(8k)})$, which is close to $o(n^{1/2})$ for $k \geqslant 3$ and improves the earlier rate $o(n^{1/3})$ attained in Hjort et al. (2008). By reviewing the proof of Theorem 3, we can see that if Z_{ij} are all bounded random variables the dimensionality p can reach $o(n^{1/2})$. We believe that $p = o(n^{1/2})$ is the best rate for the asymptotic normality of the empirical likelihood ratio with the normalizing constants p and $(2p)^{1/2}$, based on the following considerations. Lemma 4 in the Appendix implies that, when the largest eigenvalue of Σ is bounded, $||S_n - \Sigma||_{\mathrm{tr}} \to 0$ in probability if and only if $p = o(n^{1/2})$. Here $||A||_{\mathrm{tr}} = \{\mathrm{tr}(A'A)\}^{1/2}$ is the trace norm. Bai & Yin (1993) established the convergence of S_n to Σ with probability one if $p = o(n)$ under the matrix norm based on the largest eigenvalue by assuming each Z_{nl} are independent and identically distributed. However, it can be seen from our proofs in the technical report that the convergence of S_n to Σ under the trace norm is the one used in establishing various results for the empirical likelihood.

As shown by Theorems 2 and 3, when (3) is valid, the asymptotic mean and variance of the empirical likelihood ratio are respectively p and $2p$, which are known. This means that the empirical likelihood carries out internal studentizing even when p increases along with n. However, it is apparent that the internal studentization prevents p from growing faster as it brings in those higher-order terms.

The least-squares empirical likelihood is a simplified version of the empirical likelihood. The least-squares empirical likelihood ratio for μ is $q_n(\mu) = \min \sum(n\pi_i - 1)^2$ subject to $\sum_{i=1}^n \pi_i = 1$ and $\sum \pi_i(X_i - \mu) = 0$. The least-squares empirical likelihood uses $\sum(n\pi_i - 1)^2$ to approximate $-2\sum \log(n\pi_i)$. As shown in Brown & Chen (1998), the optimal weights π_i admit closed-form solutions so that

$$q_n(\mu) = n(\bar{X}-\mu)^{\mathrm{T}} H_n^{-1}(\bar{X}-\mu), \tag{11}$$

where $H_n = S_n - (\bar{X} - \mu)(\bar{X} - \mu)^{\mathrm{T}}$. Thus, $q_n(\mu)$ can be readily computed without solving the nonlinear equation (5) as for the full empirical likelihood. The least-squares empirical likelihood ratio is a first-order approximation to the full empirical likelihood ratio, and $q_n(\mu) \to \chi_p^2$ in distribution when p is fixed.

The least-squares empirical likelihood is less affected by higher dimension. In particular, if $k \geqslant 3$ in (4), then

$$(2p)^{-1/2}\{q_n(\mu) - p\} \to N(0, 1) \tag{12}$$

in distribution as $n \to \infty$ when $p = o(n^{2/3})$, which improves the rate given by Theorem 3 for the full empirical likelihood ratio $w_n(\mu)$.

To appreciate (12), we note from (11) that

$$q_n(\mu) = n(\bar{X} - \mu)^{\mathrm{T}}\Sigma^{-1}(\bar{X} - \mu) + n(\bar{X} - \mu)^{\mathrm{T}}(H_n^{-1} - \Sigma^{-1})(\bar{X} - \mu). \tag{13}$$

Then, following a similar line to the proof of Lemma 6,

$$n(\bar{X} - \mu)^{\mathrm{T}}(H_n^{-1} - \Sigma^{-1})(\bar{X} - \mu) = O_p(p^2/n) = o_p(p^{1/2}).$$

As the first term on the right-hand side of (13) is asymptotically normal with mean p and variance $2p$ as conveyed in (10), (12) is valid.

If we confine ourselves to specific distributions, faster rates for p can be established. For example, if the data are normally distributed, the least-squares empirical likelihood ratio is the Hotelling-T^2 statistic, which is shown in Bai & Saranadasa (1996) to be asymptotically normal if $p/n \to c \in [0, 1)$.

4. NUMERICAL RESULTS

We report results from a simulation study designed to evaluate the asymptotic normality of the empirical likelihood ratio. The $p \times 1$ independent and identically distributed data vectors $\{X_i\}_{i=1}^n$ were generated from a moving average model,

$$X_{ij} = Z_{ij} + \rho Z_{ij+1} \quad (i = 1, \ldots, n, \quad j = 1, \ldots, p),$$

where, for each i, the innovations $\{Z_{ij}\}_{j=1}^{p+1}$ were independent random variables with zero mean and unit variance. We considered two distributions for the innovation Z_{ij}. One is the standard normal distribution, and the other is a standardized version of a Pareto distribution with distribution function $(1 - x^{-4\cdot5})I(x \geqslant 1)$. We standardized the Pareto random variables so that they had zero mean and unit variance. As the Pareto distribution has only four finite moments, we had $k = 1$ in (4), whereas $k = \infty$ for the normally distributed innovations. In both distributions, X_i is a multivariate random vector with zero mean and covariance $\Sigma = (\sigma_{ij})_{p\times p}$, where $\sigma_{ii} = 1, \sigma_{ii\pm1} = \rho$ and $\sigma_{ij} = 0$ for $|i - j| > 1$. We set ρ to be 0·5 throughout the simulation.

To make p and n increase simultaneously, we considered two growth rates for p with respect to n: (i) $p = c_1 n^{0\cdot4}$ and (ii) $p = c_2 n^{0\cdot24}$. We chose the sample size $n = 200, 400$ and 800. By assigning $c_1 = 3, 4$ and 5 in the faster growth rate setting (i), we obtained three dimensions for each sample size, which were $p = 25, 33$ and 43 for $n = 200$; $p = 33, 44$ and 58 for $n = 400$; and $p = 42, 55$ and 72 for $n = 800$, respectively. For the slower growth rate setting (ii), to maintain a certain amount of increase between successive dimensions when n was increased, we assigned larger $c_2 = 4, 6$ and 8, which led to $p = 14, 17$ and 20 for $n = 200$; $p = 21$, 25 and 30 for $n = 400$; and $p = 29$, 34 and 40 for $n = 800$, respectively.

We carried out 500 simulations for each of the (p, n)-combinations and for each of the two innovation distributions. Figure 1 displays Q-Q plots of standardized empirical likelihood ratio

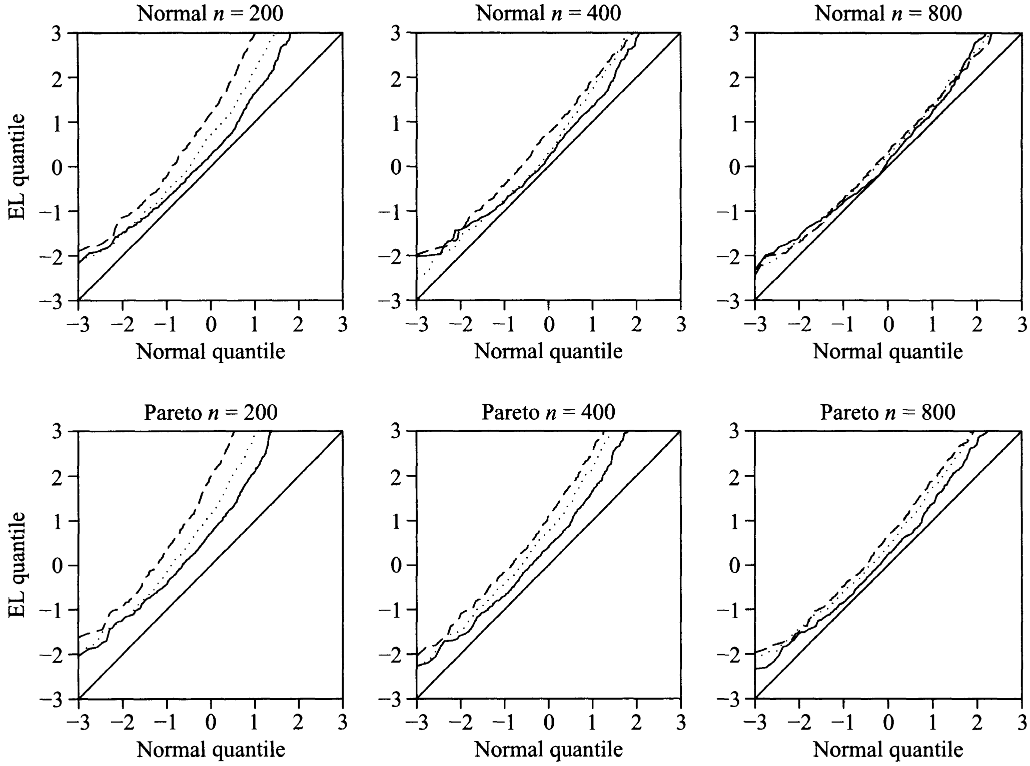

Fig. 1. Normal Q-Q plots with the faster growth rate $p = c_1 n^{0\cdot 4}$ for the normal (upper panels) and the Pareto (lower panels) innovations: $c_1 = 3$ (solid line), 4 (dotted lines) and 5 (dashed lines).

statistics for the faster growth rate (i). Those for the slower growth rate (ii) are presented in Fig. 2. As n and p were increased simultaneously, there was a general convergence of the standardized empirical likelihood ratio to $N(0, 1)$. We also observed that the convergence in Fig. 2 for the slower growth rate setting (ii) was faster than that in Fig. 1 for the faster growth rate setting. This is expected as the setting (i) ensured much higher dimensionality. The convergence for the normal innovation was faster than that for the Pareto case when $p = c_1 n^{0\cdot 4}$ in Fig. 1. This may be explained by the fact that the Pareto distribution has only finite fourth moments, which corresponds to $k = 1$, whereas the normal innovation has all moments finite. According to Theorems 2 and 3, the growth rate for p depends on the value of k: the larger the k, the higher the rate. For the lower growth rate in setting (ii), Fig. 2 shows that, there was substantial improvement in the convergence in the Q-Q plots as p was increased at the slower rate for both distributions of innovations.

It is observed that the most of the lack-of-fit in the $N(0, 1)$ Q-Q plots in Figs. 1 and 2 appeared at the lower and upper quantiles. This could be attributed to the lack-of-fit between χ_p^2 and $N(0, 1)$, as χ_p^2 may be viewed as the intermediate convergence of the empirical likelihood ratio.

To verify this, we carried out further simulations by inverting settings (i) and (ii) so that for a given dimension p, three sample sizes were generated according to (iii) $n = (p/c_1)^{1/0\cdot 4}$ and (iv) $n = (p/c_2)^{1/0\cdot 24}$, with $c_1 = 3, 4$ and 5 and $c_2 = 4, 5$ and 6, respectively. We chose $p = 35, 45$ and 55 for the setting (iii) and $p = 17, 20$ and 25 for the setting (iv). Two figures of χ_p^2-based Q-Q plots for (iii) and (iv), given in the Iowa State University technical report, show that there was a substantial improvement in the overall fit of the Q-Q plots, and that the lack-of-fit in the $N(0, 1)$-based Q-Q plots largely disappeared.

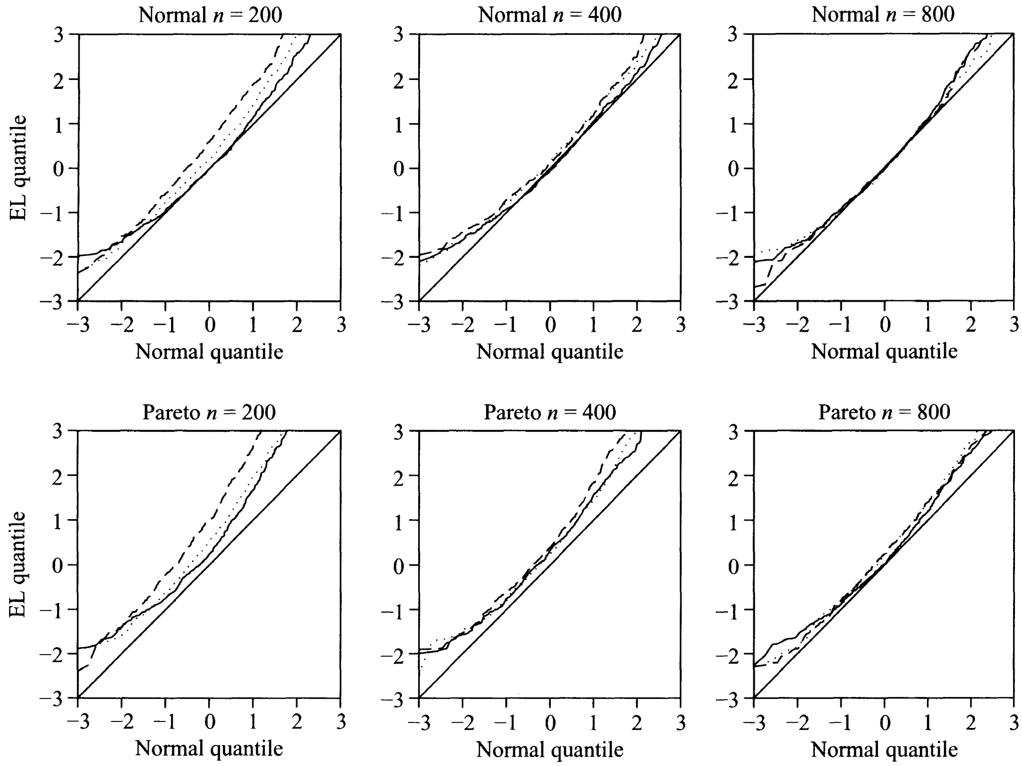

Fig. 2. Normal Q-Q plots with the slower growth rate $p = c_2 n^{0\cdot 24}$ for the normal (upper panels) and the Pareto (lower panel) innovations: $c_2 = 4$ (solid line), 6 (dotted lines) and 8 (dashed lines).

ACKNOWLEDGEMENT

The authors thank the editor, Professor D. M. Titterington, and two referees for valuable suggestions that have improved the presentation of the paper. The authors acknowledge grants from the U.S. National Science Foundation.

APPENDIX

Technical details

We first establish some lemmas.

LEMMA 1. *If $m_{4k} < \infty$ for some $k \geqslant 1$, then*

$$E(||X_i - \mu||^{2k}) = O\{\mathrm{tr}^k(\Sigma)\} \quad \textit{and} \quad \mathrm{var}(||X_i - \mu||^{2k}) = O\{\mathrm{tr}^{2k-1}(\Sigma)\gamma_p\}.$$

Proof. We only show the case of $k = 1$ since other cases are similar. It is easy to check that

$$E(||X_i - \mu||^2) = \mathrm{tr}\{E(X_i - \mu)^{\mathrm{T}}(X_i - \mu)\} = \mathrm{tr}(\Sigma) \tag{A1}$$

and

$$E(||X_i - \mu||^4) = E(||\Gamma Z_i||^4) = E\big(Z_i^{\mathrm{T}}\Gamma^{\mathrm{T}}\Gamma Z_i Z_i^{\mathrm{T}}\Gamma^{\mathrm{T}}\Gamma Z_i\big) = \mathrm{tr}\big\{\Gamma^{\mathrm{T}}\Gamma E\big(Z_i Z_i^{\mathrm{T}}\Gamma^{\mathrm{T}}\Gamma Z_i Z_i^{\mathrm{T}}\big)\big\}.$$

Write $\Gamma^{\mathrm{T}}\Gamma = (\nu_{sl})_{1\leqslant s,l\leqslant m}$. Then $Z_i Z_i^{\mathrm{T}}\Gamma^{\mathrm{T}}\Gamma Z_i Z_i^{\mathrm{T}} = (\sum_{j=1}^m \sum_{l=1}^m Z_{ik_1} Z_{il}\nu_{lj} Z_{ij} Z_{ik_2})_{1\leqslant k_1,k_2\leqslant m}$. When $k_1 = k_2 = s$,

$$E\big\{\textstyle\sum_{j=1}^m \sum_{l=1}^m Z_{ik_1} Z_{il}\nu_{lj} Z_{ij} Z_{ik_2}\big\} = \nu_{ss} E(Z_{is})^4 + \sum_{l\neq s} \nu_{ll}.$$

When $k_1 \neq k_2$, $E(\sum_{j=1}^m \sum_{l=1}^m Z_{ik_1} Z_{il} \nu_{lj} Z_{ij} Z_{ik_2}) = 2\nu_{k_1 k_2}$. Hence,

$$E(||X_i - \mu||^4) = \{m_4 - 3\} \sum_{s=1}^m \nu_{ss}^2 + \mathrm{tr}^2(\Sigma) + 2\mathrm{tr}(\Sigma^2). \tag{A2}$$

Note that $\sum_{s=1}^m \nu_{ss}^2 \leqslant \sum_{j=1}^m \sum_{s=1}^m \nu_{js}^2 = \mathrm{tr}\{(\Gamma^{\mathrm{T}}\Gamma)^2\} = \mathrm{tr}(\Sigma^2)$. This together with (A1) and (A2) implies that $\mathrm{var}(||X_i - \mu||^2) = \{m_4 - 3\} \sum_{s=1}^m \nu_{ss}^2 + 2\mathrm{tr}(\Sigma^2) = O\{\mathrm{tr}(\Sigma^2)\}$. □

LEMMA 2. *If $m_{4k} < \infty$ for some $k \geqslant 1$, then, with probability one,*

$$\max_{i=1,\ldots,n} ||X_i - \mu|| = o\left[\{\mathrm{tr}(\Sigma)\}^{-(2k-1)/(4k)} \gamma_p^{1/(4k)} n^{1/(4k)}\right] + O\{\mathrm{tr}^{1/2}(\Sigma)\}.$$

Proof. We note that

$$\max_{i=1,\ldots,n} ||X_i - \mu|| \leqslant \left\{ \max_{i=1,\ldots,n} |||X_i - \mu||^{2k} - E(||X_i - \mu||^{2k})| + E(||X_i - \mu||^{2k}) \right\}^{1/(2k)}$$

and

$$\max_{i=1,\ldots,n} [||X_i - \mu||^{2k} - E(||X_i - \mu||^{2k})\{\mathrm{var}(||X_i - \mu||^{2k})\}^{-1/2}] = o(n^{1/2})$$

with probability one as $n \to \infty$. The lemma is proved by applying Lemma A3 of Owen (1990) and Lemma 1. □

From now on, we let $Y_i = \Sigma^{-1/2}(X_i - \mu)$, $V_n = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^{\mathrm{T}}$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $D_n = V_n - I_p = (d_{sl})_{l,s=1,\ldots,p}$.

LEMMA 3. *Under the conditions of Theorem* 1, $\mathrm{tr}(D_n^2) = O_p(p^2/n)$.

Proof. We only need to show that $E\{\mathrm{tr}(D_n^2)\} = O(p^2/n)$. Note that $V_n = \Sigma^{-1/2}\Gamma S_z \Gamma^{\mathrm{T}} \Sigma^{-1/2} \tilde{\Sigma}$, where $S_z = n^{-1}\sum_{i=1}^n Z_i Z_i^{\mathrm{T}}$ and $\tilde{\Sigma} = \Gamma^{\mathrm{T}}\Sigma^{-1}\Gamma = (\tilde{\sigma}_{jl})_{j,l=1,\ldots,m}$, say. Then

$$\mathrm{tr}(D_n^2) = \mathrm{tr}(S_z \tilde{\Sigma} S_z \tilde{\Sigma}) - 2\mathrm{tr}(S_z \tilde{\Sigma}) + p \tag{A3}$$

and

$$E\{\mathrm{tr}(S_z \tilde{\Sigma})\} = E\left(\sum_{j,l=1}^m n^{-1} \sum_{i=1}^n Z_{ij} Z_{il} \tilde{\sigma}_{lj} \right) = \sum_{j,l=1}^m \delta_{jl} \tilde{\sigma}_{lj} = \sum_{j=1}^m \tilde{\sigma}_{jj} = p \tag{A4}$$

since $\mathrm{tr}(\tilde{\Sigma}) = \mathrm{tr}(I_p) = p$. By utilizing information of Z_i in (4),

$$\begin{aligned} E[\mathrm{tr}\{(S_Z \tilde{\Sigma})^2\}] &= E\left(\sum_{j,l=1}^m \sum_{l_1,l_2=1}^m n^{-2} \sum_{i_1,i_2=1}^n Z_{i_1 j} Z_{i_1 l_1} Z_{i_2 l} Z_{i_2 l_2} \tilde{\sigma}_{l_1 l} \tilde{\sigma}_{l_2 j} \right) \\ &= m_4 n^{-1} \sum_{j=1}^m \tilde{\sigma}_{jj}^2 + n^{-1} \sum_{j \neq l} \left(2\tilde{\sigma}_{jl}^2 + \tilde{\sigma}_{jj}\tilde{\sigma}_{ll}\right) + (1 - n^{-1}) \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 \\ &= \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 + (m_4 - 1)n^{-1} \sum_{j=1}^m \tilde{\sigma}_{jj}^2 + n^{-1} \sum_{j \neq l} \left(\tilde{\sigma}_{jl}^2 + \tilde{\sigma}_{jj}\tilde{\sigma}_{ll}\right). \end{aligned}$$

It is easy to check that $\sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = \mathrm{tr}(\tilde{\Sigma}^2) = p$, $\sum_{j=1}^m \tilde{\sigma}_{jj}^2 \leqslant \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = p$, $\sum_{j \neq l} \tilde{\sigma}_{jl}^2 \leqslant \sum_{j,l=1}^m \tilde{\sigma}_{jl}^2 = p$ and $|\sum_{j \neq l} \tilde{\sigma}_{jj}\tilde{\sigma}_{ll}| \leqslant (\sum_{j=1}^m \tilde{\sigma}_{jj})^2 = p^2$. These together with (A3) and (A4) imply $E\{\mathrm{tr}(D_n^2)\} = O(p^2/n)$. □

LEMMA 4. *Under condition* (4), $\max_{i=i,\ldots,p} |\gamma_i(S_n) - \gamma_i(\Sigma)| = O_p(\gamma_p p n^{-1/2})$.

Proof. Note that

$$|\gamma_i(S_n) - \gamma_i(\Sigma)|^2 \leqslant \sum_{i=1}^{p} |\gamma_i^{1/2}(S_n^2) - \gamma_i^{1/2}(\Sigma^2)|^2$$
$$= \operatorname{tr}(S_n^2) + \operatorname{tr}(\Sigma^2) - 2\sum_{i=1}^{p} \gamma_i(S_n)\gamma_i(\Sigma).$$

By Von Neumann's inequality, $\sum_{i=1}^{p} \gamma_i(S_n)\gamma_i(\Sigma) \geqslant \operatorname{tr}(S_n\Sigma)$. Hence

$$\max_{i=1,\dots,p} |\gamma_i(S_n) - \gamma_i(\Sigma)| \leqslant \{\operatorname{tr}(S_n - \Sigma)^2\}^{1/2}.$$

Now

$$\operatorname{tr}\{(S_n - \Sigma)^2\} = \operatorname{tr}(D_n \Sigma D_n \Sigma) \leqslant \gamma_p^2(\Sigma)\operatorname{tr}(D_n^2) = O_p\{\gamma_p^2(\Sigma)p^2/n\}$$

by applying Lemma 3. □

This lemma implies that all the eigenvalues of S_n converge to those of Σ uniformly at the rate of $O_p(\gamma_p p n^{-1/2})$.

Proof of Theorem 1. By (5), $\lambda \in R^p$ satisfies

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^{\mathrm{T}}(X_i - \mu)} = g(\lambda). \quad \text{(A5)}$$

Write $\lambda = \rho\theta$, where $\rho \geqslant 0$ and $||\theta|| = 1$. Hence,

$$0 = ||g(\rho\theta)|| \geqslant |\theta^{\mathrm{T}} g(\rho\theta)|$$
$$= n^{-1}\left|\theta^{\mathrm{T}}\left\{\sum_{i=1}^{n}(X_i - \mu) - \rho\sum_{i=1}^{n}\frac{(X_i - \mu)\theta^{\mathrm{T}}(X_i - \mu)}{1 + \rho\theta^{\mathrm{T}}(X_i - \mu)}\right\}\right|$$
$$\geqslant \rho\theta^{\mathrm{T}} S_n \theta\{1 + \rho \max_{i=1,\dots,n} ||X_i - \mu||\}^{-1} - n^{-1}\left|\sum_{i=1}^{n}\theta^{\mathrm{T}}(X_i - \mu)\right|.$$

Hence,

$$\rho\left\{\theta^{\mathrm{T}} S_n\theta - \max_{i=1,\dots,n} ||X_i - \mu||n^{-1}\left|\sum_{i=1}^{n}\theta^{\mathrm{T}}(X_i - \mu)\right|\right\} \leqslant n^{-1}\left|\sum_{i=1}^{n}\theta^{\mathrm{T}}(X_i - \mu)\right|.$$

Since $n^{-1}|\sum_{i=1}^{n}\theta^{\mathrm{T}}(X_i - \mu)| = O_p[\{\operatorname{tr}(\Sigma)/n\}^{1/2}]$, it follows from Lemma 2 that

$$\max_{i=1,\dots,n} ||X_i - \mu||n^{-1}\left|\sum_{i=1}^{n}\theta^{\mathrm{T}}(X_i - \mu)\right| = o_p\{\{\operatorname{tr}(\Sigma)\}^{1-1/(4k)}\gamma_p^{1/(4k)}n^{-1/2+1/(4k)} + O_p\{\operatorname{tr}(\Sigma)n^{-1/2}\} = o_p(1). \quad \text{(A6)}$$

By Lemma 4, for a positive constant C_1, $P(\theta^{\mathrm{T}} S_n\theta \geqslant \frac{1}{2}C_1) \to 1$ as $n \to \infty$. Hence $||\lambda|| = \rho = O_p[\{\operatorname{tr}(\Sigma)/n\}^{1/2}]$. □

By repeating (A6) in the proof of the above theorem and Lemma 2, we have

$$\max_{i=1,\dots,n} ||\lambda^{\mathrm{T}}(X_i - \mu)|| \leqslant ||\lambda|| \max_{i=1,\dots,n} ||X_i - \mu|| = o_p(1). \quad \text{(A7)}$$

We need the following lemmas to prove Theorem 2.

LEMMA 5. *If $p/n \to c \geqslant 0$, then $(2p)^{-1/2}\{n(\bar{X} - \mu)^{\mathrm{T}}\Sigma^{-1}(\bar{X} - \mu) - p\} \to N(0, 1)$ in distribution as $n \to \infty$.*

Proof. The proof entails applying the martingale central limit theorem (Hall & Hyde, 1980). Bai & Saranadasa (1996) used this approach to establish asymptotic normality of a two sample test statistic for high-dimensional data. □

LEMMA 6. *Under the conditions of Theorem* 2,

$$n(\bar{X}-\mu)^{\mathrm{T}}\left(S_n^{-1}-\Sigma^{-1}\right)(\bar{X}-\mu)=o_p(p^2n^{-1/2}).$$

Proof. Recall that $D_n = V_n - I_p = (d_{sl})_{1\leqslant s\leqslant p, 1\leqslant l\leqslant p}$. It follows from Lemma 3 that

$$P\Big(\max_{k_1,k_2=1,\ldots,p}|d_{k_1k_2}|>\epsilon\Big)\leqslant\sum_{k_1=1}^{p}\sum_{k_2=1}^{p}\epsilon^{-2}E\left(d_{k_1k_2}^2\right)=\epsilon^{-2}E\left\{\mathrm{tr}\left(D_n^2\right)\right\}=O(p^2/n).$$

Hence, $d_{jl}=O_p(pn^{-1/2})=o_p(1)$ uniformly in $1\leqslant j,l\leqslant p$. It is easy to check that

$$V_n^{-1}-I_p=-D_n+D_n^2+D_n^2\left(V_n^{-1}-I_p\right)$$

and

$$n(\bar{X}-\mu)^{\mathrm{T}}\left(S_n^{-1}-\Sigma^{-1}\right)(\bar{X}-\mu)=n\bar{Y}^{\mathrm{T}}\left(V_n^{-1}-I_p\right)\bar{Y}.$$

From Lemma 1, $E(||\bar{Y}||^2)=n^{-1}E(||Y_1||^2)=p/n$. Since $|\bar{Y}^{\mathrm{T}}A\bar{Y}|\leqslant||\bar{Y}||^2\{\mathrm{tr}(A^2)\}^{1/2}$ for any symmetric matrix A, it follows from Lemma 4 and the condition $p=o(n^{1/3})$ that

$$|n\bar{Y}^{\mathrm{T}}D_n\bar{Y}|\leqslant n||\bar{Y}||^2\left\{\mathrm{tr}\left(D_n^2\right)\right\}^{1/2}=O_p(p^2n^{-1/2})=o_p(p^{1/2}).$$

Similarly, $|n\bar{Y}^{\mathrm{T}}D_n^2\bar{Y}|\leqslant n||\bar{Y}||^2\mathrm{tr}(D_n^2)=O_p(p^3/n)=o_p(p^{1/2})$.

Furthermore, we note the following facts:

$$|\bar{Y}^{\mathrm{T}}D_n^3\bar{Y}|\leqslant\max_{i=1,\ldots,p}\{|\gamma_i(D_n)|\}\bar{Y}^{\mathrm{T}}D_n^2\bar{Y}=o_p\left(\bar{Y}^{\mathrm{T}}D_n^2\bar{Y}\right)$$

since $\max_{i=1,\ldots,p}\{|\gamma_i(D_n)|\}\leqslant\{\mathrm{tr}(D_n^2)\}^{1/2}\to 0$ and $\bar{Y}^{\mathrm{T}}D_n^4\bar{Y}\leqslant\gamma_p(D_n^2)\bar{Y}^{\mathrm{T}}D_n^2\bar{Y}=o_p(\bar{Y}^{\mathrm{T}}D_n^2\bar{Y})$. In general, if $p=o(n^{1/2})$, for any positive integer l,

$$\bar{Y}^{\mathrm{T}}D_n^{2+l}\bar{Y}=o_p\left(\bar{Y}^{\mathrm{T}}D_n^2\bar{Y}\right).$$

The lemma follows from summarizing the above results. □

Proof of Theorem 2. Put $W_i=\lambda^{\mathrm{T}}(X_i-\mu)$. Then (A7) implies that $\max_{i=1,\ldots,n}|W_i|=o_p(1)$. Expand equation (A5),

$$0=g(\lambda)=\bar{X}-\mu-S_n\lambda+\beta_n \tag{A8}$$

where $\beta_n=n^{-1}\sum_{i=1}^n(X_i-\mu)\frac{W_i^2}{(1+\xi_i)^3}$ and $|\xi_i|\leqslant|\lambda^{\mathrm{T}}(X_i-\mu)|$. As $\max_{i=1,\ldots,n}|W_i|=o_p(1)$, $\max_{i=1,\ldots,n}|\xi_i|=o_p(1)$ as well. Hence $\beta_n=\beta_{n1}\{1+o_p(1)\}$, where $\beta_{n1}=n^{-1}\sum_{i=1}^n(X_i-\mu)W_i^2$. Apply Theorem 1 and Lemma 2 with $k=1$, we have, if $\mathrm{tr}(\Sigma)=O(\gamma_p^{5/3}n^{1/3})$,

$$||\beta_{n1}||=\max_{i=1,\ldots,n}||X_i-\mu||||\lambda||^2O_p(\gamma_p(\Sigma))=o_p(||\lambda||). \tag{A9}$$

It follows from (A8) that

$$\lambda=S_n^{-1}(\bar{X}-\mu)+S_n^{-1}\beta_n \tag{A10}$$

and $\log(1+W_i)=W_i-W_i^2/2+W_i^3/(1+\xi_i)^4$ for some ξ_i such that $|\xi_i|\leqslant|W_i|$. Therefore,

$$\begin{aligned}w_n(\mu)&=n(\bar{X}-\mu)^{\mathrm{T}}S_n^{-1}(\bar{X}-\mu)-n\beta_nS_n^{-1}\beta_n+\tfrac{2}{3}\textstyle\sum_{i=1}^n\left\{\lambda^{\mathrm{T}}(X_i-\mu)\right\}^3(1+\xi_i)^{-4}\\&=n(\bar{X}-\mu)^{\mathrm{T}}\Sigma^{-1}(\bar{X}-\mu)+n(\bar{X}-\mu)^{\mathrm{T}}\left(S_n^{-1}-\Sigma^{-1}\right)(\bar{X}-\mu)\\&\quad-n\beta_nS_n^{-1}\beta_n+\tfrac{2}{3}R_n\{1+o_p(1)\},\end{aligned}$$

where $R_n = \sum_{i=1}^{n}\{\lambda^{\mathrm{T}}(X_i - \mu)\}^3$. By (A9), (A10) and Lemma 4,

$$\left|n\beta_n S_n^{-1}\beta_n\right| \leqslant n||\beta_n||^2/\gamma_1(S_n)$$
$$= O_p\left\{\gamma_p^2 \mathrm{tr}^3(\Sigma)n^{-1}\right\} + o_p\left\{\gamma_p^{5/2}\mathrm{tr}^{5/2}(\Sigma)n^{-1/2}\right\} = o_p(p^{1/2}).$$

We also note that $R_n = (n\lambda^{\mathrm{T}}S_n\lambda)^{1/2}(\sum_{i=1}^{n}||\lambda||^4||X_i - \mu||^4)^{1/2} = o_p(p^{1/2})$. Hence the theorem follows from Lemmas 5 and 6. □

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

BAI, Z. & SARANADASA, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6**, 311–29.

BAI, Z. D., SILVERSTEIN, J. W. & YIN, Y. Q. (1998). A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Mult. Anal.* **26**, 166–68.

BAI, Z. & YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Prob.* **21**, 1276–94.

BROWN, B. M. & CHEN, S. X. (1998). Combined and least squares empirical likelihood. *Ann. Inst. Statist. Math.* **50**, 697–714.

CHEN, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* **83**, 329–41.

CHEN, S. X. & CUI, H. J. (2006). On bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93**, 215–20.

DICICCIO, T. J., HALL, P. & ROMANO, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053–61.

FANG, K. T & ZHANG, Y. T. (1990). *Generalized Multivariate Analysis*. Beijing: Science Press.

HALL, P. G. & HYDE, C. C. (1980). *Martingale Central Limit Theory and Its Applications*. New York: Academic Press.

HALL, P. & LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.* **58**, 109–27.

HJORT, H. L., MCKEAGUE, I. W. & VAN KEILEGOM, I. (2008). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–1115.

LEDOIT, O. & WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30**, 1081–1102.

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–49.

OWEN, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90–120.

PORTNOY, S. (1984). Asymptotic behavior of M-estimations of p regression parameters with p^2/n is large. I. consistency. *Ann. Statist.* **4**, 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M estimations of p regression parameters when p^2/n is large. II. normal approximation. *Ann. Statist.* **13**, 1403–17.

QIN, J. & LAWLESS, J. (1994). Empirical likelihood and general estimating functions. *Ann. Statist.* **22**, 300–25.

SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92**, 951–56.

SCHOTT, J. R. (2007). Some high dimensional tests for a one-way MANOVA. *J. Mult. Anal.* **98**, 1825–39.

TSAO, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32**, 1215–21.

YIN, Y. Q., BAI, Z. D. & KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Prob. Theory Rel. Fields* **78**, 509–21.

[*Received July* 2007. *Revised November* 2008]